Cluster Analysis of IRIS Spectroscopic Line Profiles and SDO/AIA EUV Emission in Observations and RMHD Simulations of the Solar Atmosphere

Viacheslav Sadykov[1,2], Irina Kitiashvili[1,2], Alexander Kosovichev[1,3]

[1]NASA ARC, [2]BAERI, [3]NJIT

Spatially-resolved observations from the IRIS and SDO/AIA satellites, especially when coupled with realistic 3D RMHD simulations, are a powerful tool for analysis of processes in the solar chromosphere, transition region, and corona. However, the complexity of the data makes understanding the observations and modeling results difficult. In this work, we apply unsupervised clustering algorithms for analysis of observational and synthetic chromospheric Mg II h&k 2796Å&2803Å and transition region C II 1334Å&1335Å line profiles observed by IRIS, and extreme ultraviolet (EUV) emission observed by SDO/AIA, for various types of problems. The synthetic line profiles are computed for simulations of the quiescent solar atmosphere (using the StellarBox and RH1.5 codes). The K-Means clustering algorithm is applied, and the selection of an optimal number of clusters is supported by the average silhouette width technique. We discuss applications of the line profile clustering method to 1) visualization of computational and observational spectroscopic imaging data; 2) understanding of evolutionary trends and behavior patterns of quiet Sun emission and during solar flares; and 3) recognition of heating events and shock waves.

Description of data processing and clustering algorithms

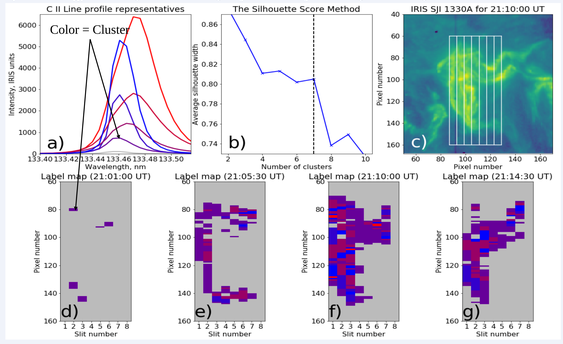
- Statistical moments of line profiles. The zeroth moment will represent the maximum or the integrated line intensity. The first (Doppler shift), second (line width), third (line asymmetry), and higher statistical moments can be computed as:

$$S_k = \sqrt[k]{\int (\lambda-\langle\lambda\rangle)^k I(\lambda)d\lambda \,/\, \int I(\lambda)d\lambda}, \qquad \langle\lambda\rangle = \int \lambda I(\lambda)d\lambda \,/\, \int I(\lambda)d\lambda$$

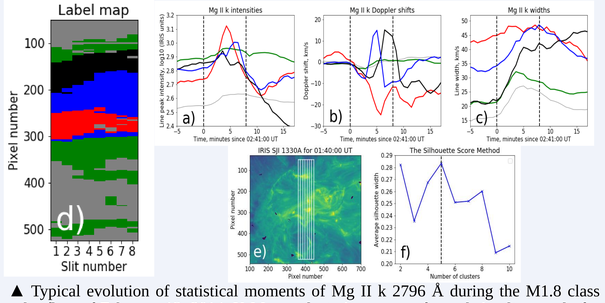
- K-Means clustering. The K-Means method takes the number of clusters as an input parameter and initially seeds the cluster centers randomly among the data points. Then each point is labeled with its nearest cluster center, the cluster centers are recomputed as the means among points of the same label, and the procedure is repeated until there are no changes in labeling.
- Average silhouette width. The silhouette is defined for data i point as $s(i) = \frac{b(i)-a(i)}{\max\{a(i),b(i)\}}$, where $a(i)$ is the average distance from point i to points of the same cluster, and $b(i)$ is the average distance from point i to points of the **closest** neighboring cluster. The average $s(i)$ across all points indicates how well the points lie within their clusters.
  - The optimal number of clusters can be estimated by maximizing $s(i)$.
  - When $s(i) < 0$, point i no longer "belongs" to its current cluster.

Identifying, from IRIS observations, the "typical" response to flare heating of the upper chromosphere and lower transition region

- The Interface Region Imaging Spectrograph (IRIS, De Pontieu et al. 2014) has observed hundreds of flares of $\geq$ C1.0, but statistical studies are hard to perform because of the complexity of imaging spectroscopy data.
- Finding "typical" responses to flare heating using unsupervised machine learning (clustering) techniques can simplify the analysis of spectroscopic data.



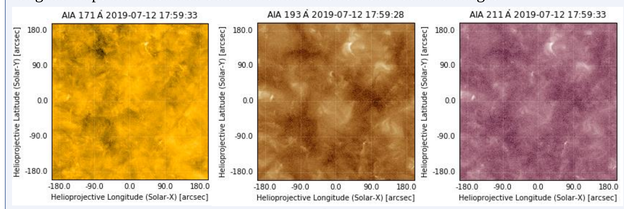
▲ An example of clustering of C II 1334.5 Å line profiles for the M1.0 flare of June 12, 2014. The clustering is done in I(λ) space. The maps of line profile representatives (panels d-g) inform by encoding the evolution of the line in time. Such clustering was previously used by Panos et al. (2018) and Sainz Dalda et al. (2019).



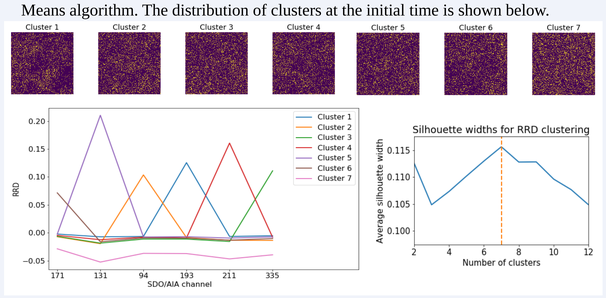
▲ Typical evolution of statistical moments of Mg II k 2796 Å during the M1.8 class solar flare of February 13, 2014. K-Means clustering was performed simultaneously for the evolutions of line intensity, Doppler shift, and line width, with equal contribution from each considered statistical moment. The color map in panel (d), supported by panels (a-c), illustrates the Mg II k line evolution during the **entire flare**. The red, blue, and black clusters are of special interest: the red cluster behaves as expected during "explosive" chromospheric evaporation, while the blue and black clusters reveal slight redshifts followed by strong blueshifts of the spectral lines.

Quantizing SDO/AIA EUV emission of the quiet Sun

We selected the 10-minute series of SDO/AIA quiet Sun observations restricted to 200 arcseconds from the disc center on July 12, 2019, 18:00 pm. Examples of the default non-aligned maps are illustrated below. Notice the offset between the images.



- Six EUV channels (all except the 304 Å channel) were aligned using the cross-correlation technique and averaged over 3x3 pixel areas, then Relative Running Differences were computed for the series: RRD(t) = I(t)/I(t-1) - 1
- The RRDs were clustered at each 12 s time moment at each pixel, and the contributions from RRDs of different channels were normalized. 7 clusters were selected for the k-Means algorithm. The distribution of clusters at the initial time is shown below.



Important: we now consider evolution of the EUV emission "quantum state" (a number ranging from 1 to 7 to indicate the cluster the point belongs to), which significantly simplifies the analysis and enhances understanding. Hereafter we will say that the RRD "is in state k" instead of "belongs to cluster k".

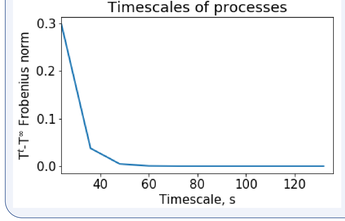
Let $P(n^{t+1}=m|n^t=k) = P(m|k)$ be the conditional probability of a point now in state k to jump to state m after 12 s. For our system, we have the very important property that $P(n^{t+1}=m|n^t=k, n^{t-1}=l) \approx P(n^{t+1}=m|n^t=k)$ for any m, k, l.

This tells us that the evolution of RRD states behaves as a Markov chain: the propagation to the next time step depends only on the current state. Markov chain dynamics is fully described by its transition matrix $T^{mk} = P(n^{t+1}=m|n^t=k) = P(m|k)$:

P(m|k)	State 1 (193Å, 1.2MK)	State 2 (94Å, 6.3MK)	State 3 (335Å, 2.5MK)	State 4 (211Å, 2.0MK)	State 5 (131Å, 0.4MK)	State 6 (171Å, 0.63MK)	State 7 (background)
State 1	**0.034**	0.130	0.133	0.151	0.134	0.141	0.146
State 2	0.149	**0.038**	0.175	0.158	0.171	0.158	0.166
State 3	0.152	0.174	**0.037**	0.159	0.169	0.156	0.165
State 4	0.163	0.144	0.147	**0.031**	0.149	0.150	0.146
State 5	0.144	0.159	0.158	0.150	**0.029**	0.149	0.153
State 6	0.149	0.149	0.148	0.152	0.150	**0.042**	0.154
State 7	0.208	0.206	0.203	0.198	0.198	0.204	**0.070**

As one can see, the temporal behavior of RRD clusters is stochastic, and the system "does not like" to be in one state for a long time (the diagonal elements of the matrix are significantly lower than the off-diagonal).

Timescales of processes. For a Markov process, the transition matrix for nt time steps is the nt-th power of the single-step transition matrix. The difference between $(T)^{nt}$ and $(T)^{\infty}$ illustrates how strongly the system "remembers" its initial state after nt transitions.



This plot shows the Frobenius norm of the matrix $(T)^{nt}$-$(T)^{\infty}$. The matrices become identical after 2-3 transitions (36-48 s), i.e., the system quickly "forgets" its initial state.

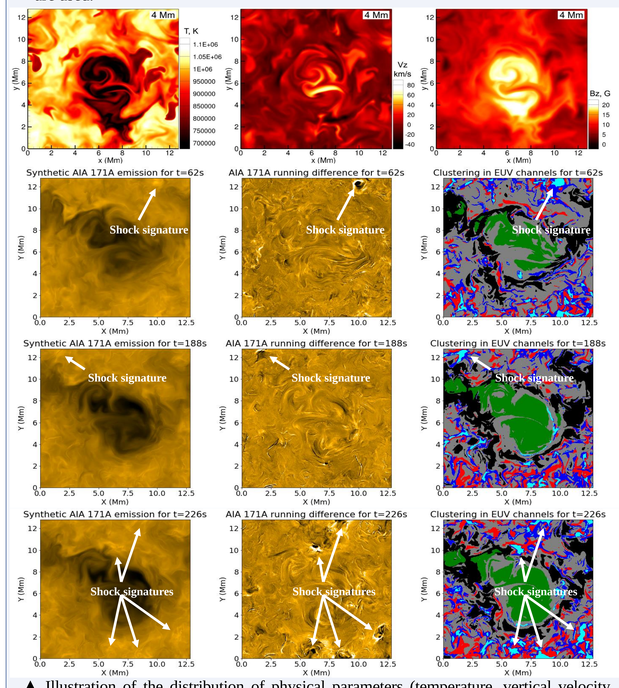
So one can write:

$P(n^{t>t0+3}=m|n^{t0}=k) = P(n^{t0}=k)$,

where $P(n^{t0}=k)$ is simply the probability to find the state k across the domain.

Recognition of shocks and heating events from synthetic AIA emissions

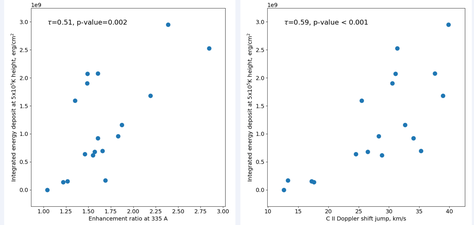
- The "StellarBox" code solves the fully compressible MHD equations with radiative transfer solved by ray-tracing and opacity binning techniques and uses a large-eddy simulation (LES) treatment of subgrid turbulent transport (Wray et al. 2015, 2018). For more details see poster SH31E-3350.
- The computational domain of 12.8 x 12.8 x 15.2 Mm includes a 10-Mm layer from the photosphere to the low corona. The horizontal resolution is 25km, and the lateral boundary conditions are periodic. 176 simulation time moments with 2s temporal cadence are analyzed.
- A Synthetic top-view AIA emission is computed for each time moment for each column separately, using SDO/AIA temperature response functions from SSW IDL. Strong impacts ("shocks" hereafter) are observed in AIA running difference images.
- K-Means clustering is performed for sparse selection of columns and snapshots for all AIA channels together. The contribution of each channel is normalized. Seven clusters are used.



▲ Illustration of the distribution of physical parameters (temperature, vertical velocity, magnetic field) at 4 Mm above the photosphere for the StellarBox setup, and the synthesized 171 Å emission, running differences, and label maps for times t=62s, t=188s, and t=226s. The cyan cluster correlates well with the shock signatures, demonstrating that shocks have distinguishable fingerprints in SDO/AIA emission.

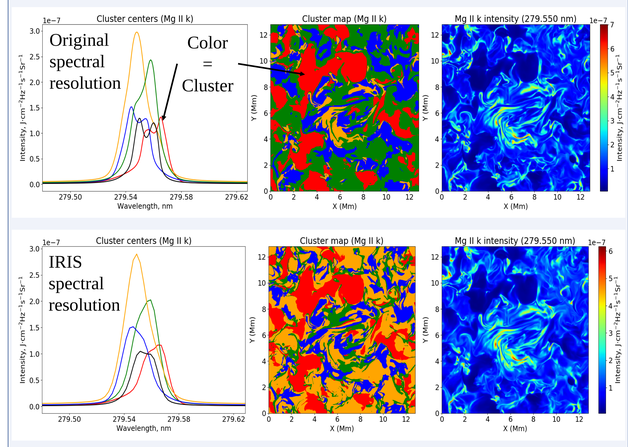
Why is it important to identify shockwaves?

Shocks are among the possible mechanisms contributing to coronal heating. Our preliminary study indicates that the properties of synthesized emission are well correlated with the energy transported by the shock to the corona. Examples of these correlations are shown below.



Clustering of synthetic Mg II spectra

- Synthetic Mg II spectra are computed for the previously-described StellarBox simulation, with 2s temporal cadence, for the last 176 simulation time moments. Non-LTE calculations are done with the RH1.5 radiative transfer code (Pereira and Uitenbroek 2015)
- We reduced the original spectral resolution to match the IRIS resolution.
- The clustering of Mg II k line profiles is done in I(λ) space. Results for the original resolution and for the IRIS-compatible spectral resolution are demonstrated below for the last time moment of the analyzed time sequence.



One can see that the reduction of spectral resolution to the IRIS-compatible one significantly smooths the lines. Asymmetry of the line peaks, as well as the line dip (central reversal), almost vanish across the domain. Clustering helps us to understand how resolution reduction affects the line profiles not simply at a single point but across the entire domain.

Plans and ideas for the future

- Recognition of typical line profiles and dynamical responses of the atmosphere to flare heating from IRIS data based on large statistics of flare events.
- Correlation of the appearance of certain line profile shapes and dynamical behavior with properties of hard X-rays (from RHESSI, Fermi GBM, and Konus-WIND) and soft X-rays (from GOES/XRS).
- Understanding the evolution of the quiet Sun EUV emission observed by SDO/AIA and spectral lines observed by IRIS based on clusterization (quantization) of the data. Analysis of the timescales of underlying processes.
- More detailed analysis of IRIS line profiles (Mg II, C II) for the considered StellarBox run and development of automatic recognition of shocks from synthesized spectra.
- Connection between simulations and observations via comparison of the statistical properties (timescales and patterns) of emission.

References

1. De Pontieu, B., Title, A.M., Lemen, J.R. et al. 2014, Solar Physics, 289, 2733
2. Panos, B., Kleint, L., Huwyler, C. et al. 2018, ApJ, 861, 62.
3. Pereira, T.M.D., and Uitenbroek, H. 2015, A&A, 574, A3
4. Sainz Dalda, A., de la Cruz Rodríguez, J., De Pontieu, B., and Gošić, M. 2019, ApJL, 875 L18.
5. Vernazza, J.E., Avrett, E.H., Loeser, R. 1981, ApJSS, 45, 635
6. Wray A.A., Bensassi K., Kitiashvili I.N. et al. 2018a, In Book: "Variability of the Sun and Sun-like Stars: from Asteroseismology to Space Weather". Eds: J.-P. Rozelot, E.S. Babaev, EDP Sciences, p.39-62.
7. Wray, A.A., Bensassi K., Kitiashvili I.N. et al. 2015, eprint arXiv:1507.07999

This research is funded by the NASA Heliophysics Supporting Research Program